Outstanding Poster Award Recipients

BioGrid Australia: Integrating Health and Research Data to Facilitate Improved Health Outcomes (#4)

Information technology is revolutionizing healthcare and medical research. Traditionally data was stored in a silo, making data extraction a nightmare of ethics approvals, comparability and logistics, however linking records about an individual and collating data from multiple sources has powerful potential to understand the causation of human disease and predict outcomes. BioGrid Australia has implemented technology and processes allowing researchers to extract data dynamically from multiple sources. BioGrid technology allows authorized access to securely held data to be transmitted in encrypted form and tracks and audits all data queries. It uses a federated design in which the data is stored at the owner's institution and only extracted with authorization. Using portal and business glossary applications, researchers explore data availability and definitions before applying for access. The analytical and query tools allow researchers to extract and analyze the data themselves. Recent advances/research findings include: (1) Cartwheel.org. This international rare tumor portal allows patients to enter the details of their illness online giving them and researchers the opportunity to participate in research projects with larger groups of patients as well as facilitating enhanced information sharing and support for patients; (2) Blood parameters as biomarkers for Brian Tumours, K Field, et al.; (3) Calculating the Rapidly Escalating Cost of Treating Cancer in Australia – Time for an Increased Focus on Prevention and Screening, Ben Tran et al.; (4) Initial impact of Australia's National Bowel Cancer Screening Program, S. Ananda et al. BioGrid is implemented in a number of states in Australia and has enabled discovery and collaborative research to be accessible via the Web while addressing security, intellectual property, and privacy.

AUTHORS AND AFFILIATIONS Robert Merriel, BA, GD Psych, GD Acc, CPA, Melbourne Health; Marienne Hibbert, B App Sci, M App Sci, GDip Epi, GDip Stats, PhD, BioGrid Australia Ltd

Experiences in selecting and installing the caBIG, bioinformatics system at a new research unit in a developing country (#14)

Introduction: Science and technological advances have created unfathomable opportunities for managing and improving health in ways that were not possible previously, yet their potential for the benefit of society has not been fully realized. Important data and information are often 'trapped' in silos among research communities that are not aware of new bodies of knowledge across disciplines, do not have access to data and/or have not fully embraced the cooperative spirit of sharing. One of the objectives of The Aga Khan University's new Clinical Research Unit (CRU) is to provide a harmonized and interoperable bioinformatics environment that will empower its research community to manage, share, examine and apply existing and new knowledge in meaningful and innovative ways across disciplines at its international campuses.
Objective: To share the experiences, processes, and resources required in selecting and installing an informatics system at a Clinical Research Unit in a developing country
Method: The process comprised of the following: System research and selection: establishing a multidisciplinary team, defining assessment parameters, documenting work and data flow, researching and evaluating software systems Planning and preparation: Team planning and prioritization, studying system documentation, establishing technical support link with caBIG team, progress review System installation: Hardware configuration; sequential installation and integration of C3PR, CAAERS, PSC, Labviewer and Open Clinica on a Windows platform. Each step included trouble shooting and local testing. Timely technical guidance of the caBIG team and the Knowledge Center were instrumental in resolving challenges.

Results: caBIG was selected on the basis of several key parameters such as functionality, flexibility, cost, standardization, continuing development, technical support and the caBIG community. System modules were installed and integrated successfully on a Windows platform over four months. A total of 892 hours were spent in this first phase constituting 21% for research and evaluation of systems, 13 % for planning and progress review and 66% for system installation, problem solving and testing.
Discussion: Successful installation in large part was attributed to the perseverance of a dedicated team and timely access to technical support. Institutions should be prepared to ride the storm of unexpected challenges during system set-up by committing adequate resources, especially dedicated staff time. The experience of the CRU has been positive, positioning it to commence the next phase of implementation involving roll out to its users.

AUTHORS AND AFFILIATIONS Rozmin Jamal, Aga Khan University; Mohsin-e Azam, Aga Khan University; Naeem Charolia, Aga Khan University; Sajida Parveen, Aga Khan University

The TRIAD Project: Adopting and Adapting caGrid for the CTSA Environment (#38)

The availability of scalable, extensible, service-oriented biomedical informatics platforms is critical to the performance of efficient, timely, and high quality research in multi-site or investigator settings. Such requirements are further emphasized when research is being conducted across the full translational spectrum involving a broad variety of information sources, stakeholders, and analytic resources. We report upon the design, initial adoption, and future plans for such an informatics platform, as part of the activities of The Ohio State University's CTSA-funded Center for Clinical and Translational Science, which is known as the Translational Informatics and Data-management (TRIAD) project. As part of the TRIAD, we have deployed and extended the caGrid middleware as our core data and knowledge-sharing platform. Specific extensions to caGrid have included: 1) the implementation of OpenMDR, an extensible knowledge management component that extends the UK's CancerGrid metadata repository (cgMDR) platform, which is an open standards and open source implementation of the ISO metadata registries standard (ISO11179-3) and allows users to define, manage, and utilize locally relevant metadata element definitions as well as access standard metadata repositories; 2) the deployment of software components capable of linking the GAARDS security system with institution-level authentication systems; and 3) the design, evaluation, and integration of software components capable of publishing or consuming information contained in common research data management systems, such as i2b2 and REDCap. At the time of this submission, TRIAD has been deployed for use at OSU, and is actively being utilized in multiple projects, including: 1) longitudinal tracking of phenotypic data for maternal-fetal dyads in support of perinatal outcomes research; 2) discovery of patient and tissue cohorts spanning numerous institutional bio-specimen repositories and an enterprise data warehouse; and 3) development of an integrative clinical trial and bio-specimen management solution for multi-center oncology research.

AUTHORS AND AFFILIATIONS Philip R.O. Payne, PhD, The Ohio State University; Justin Permar, The Ohio State University; David Ervin, The Ohio State University; Rakesh Dhaval, MS, The Ohio State University; Calixto Melean, MS The Ohio State University, ; Tara B. Borlawsky, MA, The Ohio State University

Somatic Mutations, Interoperability, Integration and Data Standards (#44)

Cancer is a genetic disease affecting approximately one in three individuals in Europe and North America. The Cancer Genome Project (CGP) was started at the Wellcome Trust Sanger Institute in 2000 to identify somatic

changes and genes implicated in cancer. As part of this project a database, the Catalogue of Somatic Mutations In Cancer (COSMIC) was developed to store mutations uncovered by the project and those curated from the scientific literature (http://www.sanger.ac.uk/genetics/CGP). In order to maximize the value of this resource we have standardized the data enabling us to integrate and interoperate with external data resources. For example, we have standardized the description of somatic mutations using HGVS nomenclature and made the data available through the Distributed Annotation System (DAS). This Web service enables external resources to integrate information from multiple distant servers and present them in a single consolidated view. This enables our somatic mutation data to be easily viewed in ensemble, allowing the user to see the genomic context of the mutation and in Pfam, to explore the protein context of these mutations. In addition to looking into the context of a mutation we provide a simple and rapid method for mining the data. This is achieved using a query orientated data management system, BioMart (http://www.sanger.ac.uk/genetics/CGP/cosmic/biomart/martview). This allows the user to rapidly mine through complex biological data using only a simple Web interface. BioMart is being used to hold the International Cancer Genome Consortium (ICGC) data. The ICGC is a project to sequence and characterize 500 tumors and normals from 50 different tumor types. This data will include expression, copy number and full genome sequence for all samples. As COSMIC and the ICGC franchise database make the data available through BioMart using similar data models the resources will be fully interoperable.

AUTHORS AND AFFILIATIONS Adam Butler, The Wellcome Trust Sanger Institute; Jon Teague, The Wellcome Trust Sanger Institute; Simon Forbes, The Wellcome Trust Sanger Institute; Dave Beare, The Wellcome Trust Sanger Institute; Nidhi Bindal, The Wellcome Trust Sanger Institute; Kenric Leung, The Wellcome Trust Sanger Institute; Rebecca Sheperd, The Wellcome Trust Sanger Institute; Sally Bamford, The Wellcome Trust Sanger Institute; Charlotte Dunham, The Wellcome Trust Sanger Institute; Andy Menzies, The Wellcome Trust Sanger Institute; Dave Jones, The Wellcome Trust Sanger Institute; Keiran Raine, The Wellcome Trust Sanger Institute; John Marshall, The Wellcome Trust Sanger Institute; Lucy Stebbings, The Wellcome Trust Sanger Institute; Chai Yin Kok, The Wellcome Trust Sanger Institute; Mingming Jai, The Wellcome Trust Sanger Institute, Peter Campbell, The Wellcome Trust Sanger Institute; P Andrew Futreal, The Wellcome Trust Sanger Institute; Michael R Stratton, The Wellcome Trust Sanger Institute

COSMIC Database and Standards for Somatic Mutations implicated in Human Cancer (#45)

Cancer is a complex disease caused by an array of genetic alterations to include point, insertion, deletion, structural rearrangement and copy number changes, the majority of which are thought to be somatic and have accumulated during the lifetime of the individual tumor. A vast amount of information has accumulated on these alterations. However, most of these data are dispersed across the scientific literature with very little standardization. This has meant using and mining the data is extremely difficult. To address this situation, we have developed the Catalogue of Somatic Mutations In Cancer (COSMIC), which was released in 2004, to curate the scientific literature for somatic mutations (http://www.sanger.ac.uk/cosmic). The majority of point mutated cancer genes have now been curated and the database has been successfully adapted to curate cancer gene fusions. Release 46 contains information on 115,737 mutations, 449,676 tumors and 8,911 publications. With the advent of Next Generation Sequencing and the anticipated increase in mutation data, COSMIC has been adapted to handle data from these technologies. In order to maintain the data quality in COSMIC and allow interoperability with other data resources the project makes use of existing data standards. For instance, the database uses HGVS mutation nomenclature to describe each mutation. The group has also been actively involved in developing data models to represent somatic mutations with other cancer databases, the NCRI Informatics Initiative and more recently the International Cancer Genome Consortium project (ICGC). If we are to maximize the data from these

new large-scale projects, it is imperative that we continue to build on the standards that have already been developed.

AUTHORS AND AFFILIATIONS Jon Teague, The Wellcome Trust Sanger Institute; Adam Butler, The Wellcome Trust Sanger Institute; Simon Forbes, The Wellcome Trust Sanger Institute Sally Bamford, The Wellcome Trust Sanger Institute; Charlotte Cole, The Wellcome Trust Sanger Institute; David Beare, The Wellcome Trust Sanger Institute; Andrew Menzies, The Wellcome Trust Sanger Institute; Chai Yin Kok, The Wellcome Trust Sanger Institute; Rebecca Shepherd, The Wellcome Trust Sanger Institute; Nidhi Bindal, The Wellcome Trust Sanger Institute; Kenric Leung, The Wellcome Trust Sanger Institute; Mingming Jia, The Wellcome Trust Sanger Institute; David Jones, The Wellcome Trust Sanger Institute; Catherine Leroy, The Wellcome Trust Sanger Institute; John Marshall, The Wellcome Trust Sanger Institute; Keiran Raine, The Wellcome Trust Sanger Institute; Lucy Stebbings, The Wellcome Trust Sanger Institute; Peter Campbell, The Wellcome Trust Sanger Institute; Michael Stratton, The Wellcome Trust Sanger Institute; Andrew Futreal, The Wellcome Trust Sanger Institute